9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

# Prosodic features of Finnish compound words

*Päivi Virkkunen, Juraj Šimko, Heini Kallio, Martti Vainio*

Department of Digital Humanities (Phonetics and Speech Synthesis Research Group), University of Helsinki, Finland

paivi.virkkunen@helsinki.fi

## Abstract

Linguistic focus is known to influence the the prosodic characteristics of syllables, (prosodic) words, as well as phrases and whole utterances. However, not much is known about the phonetic status of compound words, especially when they take part in signaling prosodic focus. In the current study we conducted a production experiment where a set of word pairs were read under three focus conditions: broad focus, and contrastive focus on either of the words in a pair. Moreover, the word pairs were produced either as a compound word or a phrase. Fundamental frequency, intensity and segmental durations were measured and compared between the different focus and phrase conditions. Results showed significant differences in the production of compound words and phrases in broad focus condition. Contrastive focus strongly affected the acoustic parameters, and those changes masked the word type differences that were found in the broad focus condition. Yet some changes in durational patterns remained also in narrow focus conditions.

**Index Terms**: speech prosody, speech production, Finnish, compound word, prosodic phrase, prosodic focus

## 1. Introduction

Finnish is an agglutinative-fusional language with high productivity. With respect to words this means that a speaker can produce almost an endless number of new words compounding existing ones.

A compound is a word consisting of two or more lexical units. Usually the first unit (modifier) specifies or limits the meaning of the latter unit (head), e.g. *kalakeitto* 'fish soup' is a soup made with fish. Phrase is a combination of two or more words belonging semantically together, e.g. *suomen kieli* 'Finnish language'. In Finnish language compounds are virtually always written together (with hyphen in some cases), seen as one word, and therefore only the latter part inflects or gets attached affixes. Phrases, in turn, are written separately. Spelling mistakes in phonemically similar compounds and phrases are common with Finnish pupils learning to write [1].

Sometimes same word pairs or phonemes can occur both as a compound word and a phrase. In these cases two ortographically differing combinations of words with different semantic meaning (e.g. *kissankello* 'harebell' and *kissan kello* 'cat's bell') are phonemically the same, and their spoken forms can be assumed to differ only by their prosodic (phrase) structure. This study focuses on the production of compounds compared to phonemically similar phrases.

The perception of compounds and noun phrases has been studied before [2], but the production and, especially the phonetic status, of compounds remain unknown. It is well known that word stress and contrastive focus affect prosodic features of words [3]. Fundamental frequency ($f_0$), intensity, and segmental durations vary according to stress. Finnish has a fixed word stress on the first syllable of a word. Consequently Finnish compound words, as one lexical unit, have a primary stress on the first syllable. Phrases, in turn, have primary stress on the first syllable on each word they contain [4]. Thus it can be presumed that compounds and phrases differ in their prosodic structure. Polysyllabic shortening may also be used in speech production to signal the number of syllables in words or larger units, e.g. phrases [5] and thus to distinguish different word types. Many studies of polysyllabic shortening focus mainly on the stressed syllable of the word of which the number of syllables is varied but the polysyllabic shortening can be manifested on the unstressed syllables, too [6]. It can also be assumed to have an effect on speech production and utterance planning so that the patterns induced by the word stress remain in different focus conditions.

The production of focus in Finnish follows a global pattern with regard to prosodic features [7]. As in many other languages, prosodic focus is not localized to the prominent word only but affects larger part of or the whole utterance. The narrow focus is usually produced by increasing the prominence of the focused part of the sentence. This can be done by increasing the $f_0$, intensity and/or duration of the specific word of interest. We investigate, whether the prosodic patterns induced by the word stress in compounds and phrases remain in different focus conditions.

Here we present a study which examined the production of Finnish compound words in order to find out, whether speakers produce them differently to denote their semantic difference. First, the production of the compound words was studied in the broad focus condition by comparing them to the two-word phrases that consisted of the same words. The differences in the patterns of production and acoustical features between word types (compound word and phrase) was examined. Second, this study was widened to examine whether contrastive focus had an effect to the production of the two word types.

## 2. Materials and Methods

Ten Finnish compound words were chosen on the grounds that they have a phonemically matching but semantically differing counterpart in a two-word phrase. Thus, prosody is supposingly the only differentiating factor in their production. These words formed minimal pairs, e.g. *kissankello* 'harebell flower' and *kissan kello* 'cat's bell', and *märkäpuku* 'wetsuit' and *märkä puku* 'wet suit'. The chosen words were embedded in identical carrier sentences. The participants were given written question-answer pairs (see examples in Table 1).

Participants read the answers aloud in three focus conditions: broad focus (later referred as BR), narrow focus on the first part of the compound or the two-word phrase (N1) and narrow focus on the second part of the compound or the two-word phrase (N2). The desired focus condition was signified by a


10.21437/SpeechProsody.2018-177

Table 1: *Examples of the sentences used in the present study with English translation. BR = broad focus, N1 = narrow focus on the first word (W1), N2 = narrow focus on the second word (W2), C = compound word, P = phrase.*

| Focus | Finnish sentence | English translation |
|---|---|---|
| BR_C | Mitä miehellä on yllään? – Hänellä on märkäpuku. | What is that man wearing? – He is wearing a wetsuit. |
| BR_P | Mitä miehellä on yllään? – Hänellä on märkä puku. | What is that man wearing? – He is wearing a wet suit. |
| N1_C | Onko tuolla sinikello? – Ei, vaan sehän on *kissan*kello. | Is that a bluebell? No, that is a *hare*bell. |
| N1_P | Onko tuolla koiran kello? – Ei, vaan sehän on *kissan* kello. | Is that a dog's bell? – No, that is a *cat's* bell. |
| N2_C | Onko tuolla villikissa?– Ei, vaan sehän on villi*sika*. | Is that a wildcat? – No, that is a wild *boar*. |
| N2_P | Onko tuolla villi lammas? – Ei, vaan sehän on villi *sika*. | Is that a wild sheep? – No, that is a wild *pig*. |

leading question and the given word was marked with italics in the text.

### 2.1. Participants and procedure

Twenty-nine native speakers of Finnish participated in the experiment. The participants (15 male, 14 female) were between 20 and 72 years of age (median 28). None reported any hearing or speech production problems. All of the speakers spoke with a neutral Helsinki area dialect. Participants gave their written consent to participate to this experiment. The instructions were presented in written form and explained orally. Words were introduced to the participants with picture pairs before recording. The speakers were instructed to speak lively and briskly. Recordings were done in a sound-treated recording studio at the Institute of Speech Sciences at the University of Helsinki using a high quality headset condenser microphone (DPA d:fine™). The sound was stored on a computer hard drive using a high quality AD converter. Each participant read 60 sentences (10 words, 2 word types, 3 focus types) in a randomized order. If the speaker made a mistake, the sentence was repeated to accomplish a correct rendering.

### 2.2. Data and statistical analysis

A total of 1740 sentences was recorded. The utterance sized recordings were segmented and labeled with Praat (version 5.4.04). Labeling was done manually for words (W0, W1 and W2) and syllables (S1, S2, S3 and S4), and vowels (V1, V2, V3 and V4) of each target word (W1 and W2). Carrier sentence (W0) was not examined. Fundamental frequency (Hz), intensity (dB), and duration (s) were calculated from each labeled part. In this study we analyzed only the syllables S1 and S3 (the first syllables of W1 and W2) and V1 and V3 (vowels of syllables S1 and S3).

To evaluate the effects of word type and focus on these acoustic measures we used mixed effect models with the following nine measures as dependent variables: $f_0$ maxima for the first and the third vowel (**V1PMa**, **V3PMa**), intensity maxima for the first and the third vowel (**V1InMa**, **V3InMa**), durations of the first and the third syllable (**S1Du**, **S3Du**), the $f_0$ and duration ratios (**V3PMa/V1PMa**, **S3Du/S1Du**) and the difference between (logarithmic) intensity values (**S3Du-S1Du**). The word type (compound or phrase) and focus type (BR, N1 or N2) were used as fixed factors, with interaction. The intercepts for speakers and words were used as random factors. The models were fitted using `lmer` function of `lme4` package in R.

Subsequently, the statistical models were analyzed using a Tukey HSD multiple comparison technique (`glht` function of `multcomp` package of R) evaluating the statistical significance of relevant differences between various estimates.

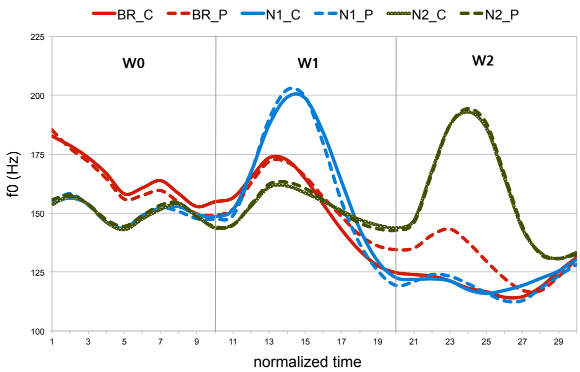


Figure 1: *Time normalized $f_0$ mean values divided by focus condition (BR, N1, N2) and word type (C=compound, P=phrase).*

## 3. Results

### 3.1. Fundamental frequency

Figure 1 shows the time normalized $f_0$ mean curves divided by focus and word type. The difference in $f_0$ between the word types in the BR-focus is clearly visible in the beginning of the second word (W2). The differences between the word types in the narrow focus conditions are much less pronounced[1].

Table 2: *Estimate $f_0$ values (Hz) of the first and third vowel of all word and focus types. V1PMa = pitch maximum of the first vowel, V3PMa = pitch maximum of the third vowel, V3PMa/V1PMa = ratio of the previous two*

| | V1PMa | V3PMa | V3PMa/V1PMa |
|---|---|---|---|
| BR_C | 180.0 | 131.4 | 0.7 |
| BR_P | 178.5 | 149.7 | 0.9 |
| N1_C | 212.4 | 127.5 | 0.6 |
| N1_P | 215.9 | 134.4 | 0.6 |
| N2_C | 167.6 | 204.3 | 1.2 |
| N2_P | 169.3 | 208.3 | 1.2 |

Table 2 contains the model-generated estimate values of $f_0$ maxima of the first (V1) and the third (V3) vowel as well as the ratio of the two; see also Fig. 2. In broad focus, for V3, the estimate values were 131.4 Hz for compound words and 149.7 Hz for phrases.The difference is statistically significant

[1]The pitch differences between the broad and narrow focus conditions in the beginning of the sentence (W0) are caused by the different sentence structures of the carrier sentences and are not analyzed.

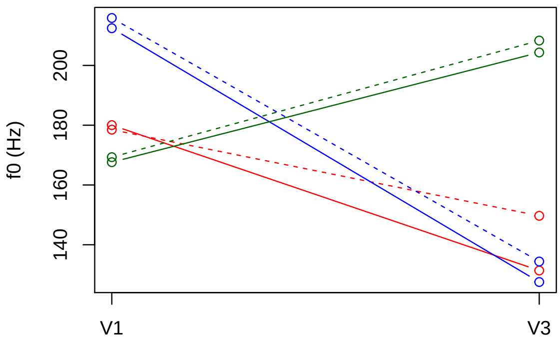


Figure 2: *The interaction plot of $f_0$ maxima of V1 and V3. Broad focus estimates are in red, N1 estimates in blue and N2 estimates in green; compounds in full and phrases in dashed lines.*

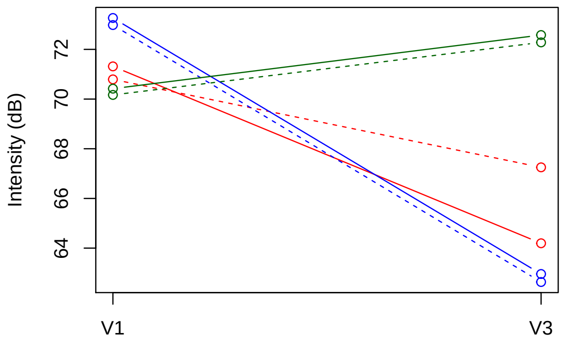


Figure 3: *The interaction plot of intensities of V1 and V3. Broad focus estimates are in red, N1 estimates in blue and N2 estimates in green; compounds in full and phrases in dashed lines.*

($t = 15.260,\ p < 0.001$). For V1, the difference between the estimates is not statistically significant ($t = -0.779,\ p = 0.94$). The ratio of V3 and V1 pitch maximas was 0.7 for compound words and 0.9 for phrases; this difference was significant ($t = 6.060,\ p < 0.001$).

In N1-condition, the V1 pitch increased significantly ($t = 16.622,\ p < 0.001$ and $t = 19.19,\ p < 0.001$ for compounds and phrases, respectively) because of the contrastive focus. The $f_0$ maxima in V1 were not significantly different between the word types in this condition ($t = 1.757,\ p = 0.364$). The unfocused V3 pitch did not change significantly for compound word, but decreased significantly for phrase ($t = -4.867,\ p < 0.001$).

Similarly, in N2-condition, the $f_0$ maximum of the focused V3 increased significantly ($t = 23.844,\ p < 0.001$ and $t = 19.798,\ p < 0.001$ for compounds and phrases, respectively); the difference in V3 pitch level was again not significantly different between the word types ($t = 1.360,\ p = 0.634$). Compared to the broad focus, the $f_0$ maximum for the unfocused V1 decreased significantly for both word types ($t = -6.352$ and $-4.721,\ p < 0.001$ for compounds and phrases, respectively); the difference between the word types was not significant ($t = 0.826,\ p = 0.917$).

### 3.2. Intensity

Table 3: *The estimate intensity values of V1 and V3 (dB) and the difference of previous two (negative values indicate decrease, positive increase). V1InMa = intensity maximum of the first vowel, V3InMa = intensity maximum of the third vowel.*

| | V1InMa | V3InMa | V3InMa-V1InMa |
|---|---|---|---|
| BR_C | 71.3 | 64.2 | -7.1 |
| BR_P | 70.8 | 67.3 | -3.5 |
| N1_C | 73.3 | 63.0 | -10.3 |
| N1_P | 73.0 | 62.6 | -10.3 |
| N2_C | 70.4 | 72.6 | 2.2 |
| N2_P | 70.2 | 72.3 | 2.1 |

Table 3 (Fig. 3) contain the model-generated estimate values of intensity of the first (V1) and the third (V3) vowel as well as the ratio of the two. The results show that in the broad focus condition the intensity maximum value of the V3 was lower than the one of V1. The difference between the vowels was -7.1 dB and -3.5 dB for compound words and phrases, respectively. The intensity decrease for the phrases was significantly lower than for the compounds ($t = 12.280,\ p < 0.001$)

In N1-condition, the V1 intensity increased significantly compared to the broad focus condition ($t = 10.147$ and $t = 11.391,\ p < 0.001$ for compounds and phrases, respectively). The intensity maxima in V1 were not significantly different between the word types in this condition ($t = -1.48,\ p = 0.550$). The unfocused V3 intensity decreased significantly for both word types, compared to the broad focus condition, ($t = -5.259$ and $-19.624,\ p < 0.001$); the difference in intensity of this vowel between the word types was not significant ($t = -1.371,\ p = 0.626$)

In N2-condition, the V3 intensity increased significantly compared to the broad focus for both word types ($t = 35.633$ and $21.364,\ p < 0.001$ for compounds and phrases, respectively). The intensity values were 72.6 dB (compounds) and 72.3 dB (phrases), the difference is not significant ($t = -1.256,\ p = 0.705$). The unfocused V1 intensity decreased significantly, compared to the broad focus values ($t = -4.671$ and $t = -3.259,\ p < 0.001$); again, the intensity difference between the word types was not significant ($t = -1.313,\ p = 0.666$).

### 3.3. Duration

Table 4: *Syllable durations (s). S1Du = duration of the S1 (first syllable of W1), S3Du = duration of the S3 (first syllable of the W2), S3Du/D1Du = the ratio of the previous two.*

| | S1Du | S3Du | S3Du/S1Du |
|---|---|---|---|
| BR_C | 0.212 | 0.226 | 1.126 |
| BR_P | 0.239 | 0.291 | 1.284 |
| N1_C | 0.283 | 0.255 | 0.958 |
| N1_P | 0.320 | 0.294 | 0.983 |
| N2_C | 0.226 | 0.339 | 1.594 |
| N2_P | 0.235 | 0.376 | 1.698 |

Table 4 contains the model-generated estimates of duration of the first and the third vowel as well as the ratio of the previous two; see also Fig. 4.

In broad focus condition the duration of both syllables was significantly longer in phrases compared to the compound

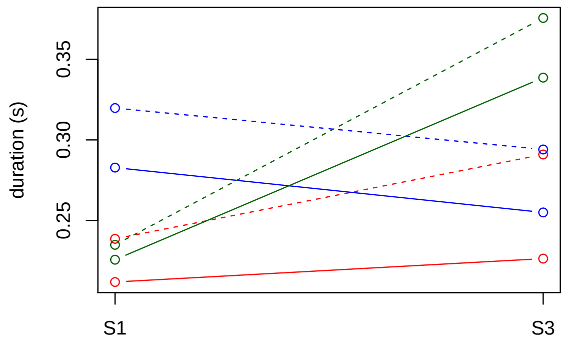


Figure 4: *Interaction plot of durations of S1 and S3. Broad focus estimates are in red, N1 estimates in blue and N2 estimates in green; compounds in full and phrases in dashed lines.*

words ($t = 7.242$ and $t = 15.260$, $p < 0.001$, for S1 and S3, respectively). The ratio of durations was significantly greater for phrases than for compounds ($t = 7.125$, $p < 0.001$).

Contrastive focus affected the syllable durations. Under the narrow focus conditions focused syllables (S1 in N1 and S3 in N2) were significantly longer than their unfocused counterparts in broad focus, both for compounds ($t = 19.188$ and $t = 26.521, p < 0.001$ for N1 and N2, respectively) as well as for phrases ($t = 21.936$ and $t = 19.981, p < 0.001$ for N1 and N2, respectively). Also, the focused syllable (S1 in N1 and S3 in N2) was significantly longer in phases compared to the compounds ($t = 9.990$ and $t = 8.730$, $p < 0.001$ for N1 and N2, respectively).

Regarding unfocused syllables, the durations of S3 in N1 and S1 in N2 increased significantly in compound words compared to the broad focus ($t = 6.767$, $p < 0.001$ and $3.736$, $p = 0.001$ for N1 and N2, respectively). For phrases, however, the difference between broad and narrow focus was not significant ($t = 0.735$, $p = 0.956$ and $t = -1.010$, $p = 0.852$ for S1 in BR vs. N2 and S3 in BR vs. N1, respectively). For N1 the unfocused syllable S3 was significantly longer in phrases than in compounds ($t = 9.228, p < 0.001$), but for N2 the difference in duration of the unfocused syllable S1 between the two word types was not significant ($t = 2.493, p = 0.075$).

## 4. Discussion

The present study examined the production of the Finnish compound words, a topic that has not been studied in this way before. The results revealed new information about the patterns speakers use when producing compound words in Finnish.

Context has a significant influence on how a listener interprets the message, but speakers can also use acoustical means to distinguish compound words from phrases in speech. The study was performed in three different focus conditions. The results showed that in the broad focus condition compound words and phrases were produced differently from each other: while $f_0$ and intensity values of V3 were decreased compared to the V1 in both word types, the decrease was significantly smaller in phrases. This can be assumed to be caused by word stress in the beginning of the second word of the phrase; speakers treated compound words as one word with one primary stress on the first syllable, while phrases as two words with individual primary stresses. The emphasis made by the contrastive focus affected the acoustic parameters in a predicted way; the overall shapes followed the ones described in [7] including post-focus compression as described in [8, 9]. In terms of $f_0$ and intensity, the increases in focused and decreases in unfocused portions actually mask the differences between the word types manifested in the broad focus condition. This is clearly illustrated in Figs. 2 and 3.

Yet some durational patterns differentiating the word types remained also in narrow focus conditions. First, the durations of the analyzed syllables were significantly longer in phrases than in compounds (except the unfocused S1 in N2 condition). Second, as expected, durations of the syllables in the focused words increased compared to the broad focus condition. Interestingly, in compound words, narrow focus leads to an increase of the duration of the unfocused syllables, but this was not the case for phrases where the duration of the syllables in unfocused words did not change compared to the broad focus condition.

The phenomenon of polysyllabic shortening suggests that speakers plan the duration of elements on the basis of the number of elements within a larger constituent [10, p. 244]; the duration of the segments decreases as their number in the word increases [11]. Also in this study syllables in compounds (i.e., longer words) were shorter than syllables in phrasal words (i.e., shorter words). Our results are also in line with durational modelling speech synthesis, e.g. Klatt's [12] [13, pp. 289-290] rule-based model for text-to-speech synthesis, which proposes rules for determination of segmental durations. The model assumes that each phonetic segment type has an inherent duration and that duration is altered based on the segment location within the word.

Recently, we replicated the experiment using pseudo words. Preliminary results show that when the words did not carry any lexical information, the speakers produced compound words and phrases the same way and no prosodic differences were found. This suggests that semantic meaning of the words, especially when making the difference between the phonemically identical words as compounds and phrases in this study, is crucial.

Further study is needed to find out whether listeners can distinguish between different word types. We can assume that distinguishing compound words and phrases in the broad focus condition should be relatively straightforward, but whether the durational patterns could help make the distinction between the word types in narrow focus conditions needs to be determined.

## 5. Conclusions

The results of the present study showed that the changes in $f_0$ and intensity that the word type created in the broad focus condition, were masked by the sentence stress in both narrow focus condition. However there were differences in the syllable duration also in narrow focus conditions. This indicates that the speaker can use syllable and word duration to signal the word type.

## 6. Acknowledgements

We would like to thank Michael O'Dell and Antti Suni for their valuable comments on this manuscript.

# 7. References


[1] P.-R. Stenberg, "Yhdyssanatajun kehittyminen ja lingvistinen tiedonrakentelu peruskoulusta yliopistoon," Master's thesis, University of Eastern Finland, 2010.

[2] A. Arnhold, "Finnish prosody: Studies in intonation and phrasing," Ph.D. dissertation, Univ.-Bibliothek Frankfurt am Main, 2013.

[3] P. Lieberman, "Some acoustic correlates of word stress in american english," *The Journal of the Acoustical Society of America*, vol. 32, no. 4, pp. 451–454, 1960.

[4] V. Koivisto, *Suomen sanojen rakenne*. Suomalaisen Kirjallisuuden Seura, 2013.

[5] L. White and A. E. Turk, "English words on the procrustean bed: Polysyllabic shortening reconsidered," *Journal of Phonetics*, vol. 38, no. 3, pp. 459–471, 2010.

[6] A. Windmann, J. Šimko, and P. Wagner, "Optimization-based modeling of speech timing," *Speech Communication*, vol. 74, pp. 76–92, 2015.

[7] M. Vainio and J. Järvikivi, "Focus in production: Tonal shape, intensity and word order," *The Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. EL55–EL61, 2007.

[8] Y. Xu, S.-W. Chen, and B. Wang, "Prosodic focus with and without post-focus compression: A typological divide within the same language family?" 2012.

[9] Y. Xu, "Effects of tone and focus on the formation and alignment of f0contours," *Journal of Phonetics*, vol. 27, no. 1, pp. 55–105, 1999.

[10] A. Turk, "The temporal implementation of prosodic structure," in *The Oxford handbook of Laboratory Phonology*, A. Cohn, C. Fougeron, and M. Huffman, Eds. Oxford: Oxford University Press, 2012, pp. 242–253.

[11] I. Lehiste, "The timing of utterances and linguistic boundaries," *The Journal of the Acoustical Society of America*, vol. 51, no. 6B, pp. 2018–2024, 1972.

[12] D. H. Klatt, "Review of text-to-speech conversion for english," *The Journal of the Acoustical Society of America*, vol. 82, no. 3, pp. 737–793, 1987.

[13] N. Campbell, "Timing in speech: A multi-level process," in *Prosody: Theory and Experiment*, M. Horne, Ed. Dordrecht: Kluwer Academic Publishers, 2000, pp. 281–334.